

REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA, 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

| | | | |
|---|------------------|---|---|
| 1. REPORT DATE (DD-MM-YYYY) | 2. REPORT TYPE | | 3. DATES COVERED (From - To) |
| | Technical Report | | - |
| 4. TITLE AND SUBTITLE IU Progress Report January 2014 Detecting Early Signatures of Persuasion in Information Cascades | | 5a. CONTRACT NUMBER W911NF-12-1-0037 | |
| | | 5b. GRANT NUMBER | |
| | | 5c. PROGRAM ELEMENT NUMBER | |
| 6. AUTHORS A. Flammini, F. Menczer, Q. Mei, S. Mailinchik | | 5d. PROJECT NUMBER | |
| | | 5e. TASK NUMBER | |
| | | 5f. WORK UNIT NUMBER | |
| 7. PERFORMING ORGANIZATION NAMES AND ADDRESSES Indiana University at Bloomington Trustees of Indiana University 509 E 3RD ST Bloomington, IN 47401 -3654 | | | 8. PERFORMING ORGANIZATION REPORT NUMBER |
| 9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES) U.S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709-2211 | | | 10. SPONSOR/MONITOR'S ACRONYM(S) ARO |
| | | | 11. SPONSOR/MONITOR'S REPORT NUMBER(S) 61766-NS-DRP.35 |
| 12. DISTRIBUTION AVAILABILITY STATEMENT Approved for public release; distribution is unlimited. | | | |
| 13. SUPPLEMENTARY NOTES The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation. | | | |
| 14. ABSTRACT Our work has focused on 1. user activity modeling and identification of roles during social movements; and, 2. anomaly detection framework for large-scale stream data analysis 3. early stage rumor detection. | | | |
| 15. SUBJECT TERMS user modeling, anomaly detection, rumor detection | | | |
| 16. SECURITY CLASSIFICATION OF: a. REPORT UU | | 17. LIMITATION OF ABSTRACT c. THIS PAGE UU | 15. NUMBER OF PAGES |
| | | | 19a. NAME OF RESPONSIBLE PERSON Alessandro Flammini |
| | | | 19b. TELEPHONE NUMBER 812-856-1830 |

Report Title

IU Progress Report January 2014

ABSTRACT

Our work has focused on

1. user activity modeling and identification of roles during social movements; and,
2. anomaly detection framework for large-scale stream data analysis
3. early stage rumor detection.

DARPA SMISC Project:

DESPIC: Detecting Early Signatures of Persuasion in Information Cascades

Teams:

Indiana University: A. Flammini (PI) and F. Menczer

University of Michigan: Qiaozhu Mei

Lockheed Martin Advanced Technology Laboratories (ATL): S. Malinchik

Progress Report – January 2014

IU: During January 2014 the Indiana University team worked on two aspect of the SMISC project: (i) user activity modeling and identification of roles during social movements; and, (ii) anomaly detection framework for large-scale stream data analysis. We also revised a manuscript about classification of promoted and organic trends and submitted it to ICWSM 2014.

Regarding point (i), our team is collecting sample datasets from the Twitter stream for social movements and other events with large number of discussion participants such as Olympic games, Oscar awards, etc. We are studying distribution of users and their roles (in the popularity-productivity space), comparing similar and distinct events to characterize universal patterns of user behavior during such events. Temporal dynamics of users in the role space and mechanisms behind users adoption of new roles have also been under investigation.

Related to point (ii), we are working on a system for detecting anomalies in large-scale stream data and to identify tweets with such anomalies. In this system, we consider computational limitations due to massive data streams to propose a scalable and efficient solution. The temporal evolution of users' and tweets' meta-data is tracked; by using heuristic techniques comparing current data with historical ones the system identifies anomalous patterns. In order to test the performance of this system we gathered 4 test cases: (1) Boston marathon bombing on April 15, 2013, (2) hurricane Sandy on October 29, 2012, (3) death of Whitney Houston on February 11, 2012, and (4) Connecticut school attack tragedy on December 14, 2012. Benchmarks will be carried out on these ad hoc test-bed scenarios in the following weeks.

From the perspective of the new infrastructure development, the IU team wrapped up the applications based on IndexedHBase, and documented the instructions about how to use it for future collaboration with other teams. We started the performance evaluation of our new system with ad hoc test cases including the analysis of one full year of Twitter stream (2012). The first benchmark consists of measuring search performance spanning this year of data. The processing was carried out on 53 nodes from the Stampede cluster of XSEDE. A Java application takes one compressed JSON file (one day of Twitter data) as input and launches multiple threads to search for tweets related to a list of text keywords or a regular expression of interest. Shell scripts are created to run multiple Java instances to simultaneously process tens to hundreds of files on several nodes. The whole year's files are split into two batches, one for 2012-01 to 2012-06, and the other for 2012-07 to 2012-

12. Overall, it took 1040 seconds to finish searching for the first batch, and 1440 seconds for the second batch. This represents about a 60-fold speed improvement with respect to the previous architecture not based on HBase. The processed result files are then stored on a data capacitor so that other members can access them. Another function of the Java application and scripts is to generate the daily frequency of all hashtags across the whole year.

Concluding, on January 31, 2014 the three teams (IU, Michigan and ATL) held a teleconference to discuss about the data challenge planned by the SMISC Data Work Group. Members from the IU team proposed this challenge to all SMISC participants. During the teleconference, details of the proposal were discussed together with possible strategies to approach this challenge. Further discussion will be required in next two weeks period. In this data challenge, the Data Work Group will provide a dataset of Twitter stream consisting of real conversations and artificially created content such as fake campaigns and rumors. Performance of each team will be evaluated by their ability to detect social bots on such simulated stream.

UM: In January, University of Michigan team has been working on the project of early stage rumor detection. The rumors we want to detect are defined as controversial and fact-checkable statements. They may discuss different topics including politics, celebrities, news and unpredicted events, etc. Spreading unconfirmed controversial information may cause potential damage to various areas such as economy and public safety. To achieve early stage rumor detection, we try to capture signal that appears before the spreading of rumors turns uncontrollable, i.e., the suspicious/uncertain attitudes and the information needs on the rumors from early exposed users. Based on this intuition, in the past several months, we designed an approach with three components for early stage rumor detection.

Our approach takes tweet stream as input and outputs potential rumors. The tweet stream will be processed in first module, which is a question & correction filter. This module will capture tweets that are questions with people's suspicion. Then we do statement detection to analyze the content of these tweets and cluster them into statements. At last, we evaluate each statement, extract features such as the popularity, level of controversy, etc. to decide whether the group is a potential rumor or not. We formalized our understanding of rumor in a codebook and trained several human annotators to help us evaluate our method. Our goal of the agreement of annotators measured by Kappa score is 0.7 and the accuracy of detecting rumor is 0.2. In January, we improved the three components of our method and trained our annotators to better understand the codebook.

To better capture the signal of users' information needs caused by uncertainty on the rumor, we collected more patterns of how people ask verification questions or post corrections about rumors based on statistical analysis. We obtained a dataset with labeled verification questions and corrections on rumors. It is used in a previous paper [Qazvinian 2011]. It contains 10,417 tweets with 3,423 tweets labeled as verification questions and corrections. We extracted ngrams from the dataset and calculated Chi-square score on each feature. We manually selected some top-ranked and content-irrelevant ngrams, such as

"unconfirmed", "debunk" or "is it real", etc., and added them to the patterns used in our question & correction filter.

We also rewrote our clustering algorithm, tuned some parameters such as minimum cluster size to improve the effectiveness and efficiency in our statement detection. Currently it can cluster more than 150,000 tweets in one run. We implemented a stream-based approach that runs our method time by time so that it will not only generate new statements, but also update old statements. We also improved our statement ranking algorithm by adopting a more efficient retrieval algorithm.

In the past month, we have been training our annotators to understand the codebook and help us label the results we generated. During the training, we discussed with them on what they didn't agree. We added more examples and resolved ambiguous explanations in our codebook. At the end of January, the inter rater reliability (Kappa score) achieved higher than 0.7. We then started to use their labeled results to evaluate our method.

In the next month, we will continue working on improving our approach evaluated by accuracy of detecting rumors. Specifically, we will train classifiers to help us decide the rank of output statements, so that statement with higher probability to be a rumor will be ranked at top. We will also compare our method with several baseline methods on different datasets. And we will think of ways to evaluate whether our method can detect rumor in the early stage.

ATL: During the past month, ATL team started working on analysis of new Twitter data set provided by Indiana University. The updated system of IU generates five different classes of features: network structure and diffusion patterns, content and language, sentiment, timing, and user features based on meta-data. The total number of features is 423. Our efforts are focused this time not only on the problem of detecting efficiently promoted content but also on identifying most informative and discriminative 5-10 features allowing to improve significantly detection performance in real time.